Abstract

Through Sentiment Analysis System it is possible to know what kind of information is there in a text. For example, one can identify if the text about a particular product, political view, sport, entertainment, education, politics, etc. or not. It is also possible to further categorize text in positive, negative or neutral. So, through proper Sentiment Analysis, the current technology would go to another step. There are so many works on Sentiment Analysis that have been done already in different languages. But due to lack of data, the work on Sentiment Analysis on Bangla Text is very limited. Because word categorization accuracy depends heavily on the size of the text corpus used to derive the inter-word statistics. So, it was planned to develop an automated corpus generation system that traverses the Web collecting text and stores them under the defined category. This flexible scheme can produce very large general-purpose corpora or particular samples of domain-specific text.
2. Xiao, R. Corpus Creation.
9. Oliver, A. Automatic creation of WordNets from parallel corpora
10. Pavel Kr´al, P., and Cerisara, C. Automatic Dialog Act Corpus creation From Web Pages

**Index Terms**

Computer Science

Information Sciences

**Keywords**

Corpus, Autonomous Corpus, Autonomous Bengali Corpus, ZIPF Law.